

Figure 1

Schematic layout of the arrangement of the genetic locus encoding the signal peptide precursor, the histidine kinase and the response regulator. Note that this arrangement is different from other loci in related streptococci for the following reasons: a) The *comC* gene is transcribed from its own promoter alone, unlike the genes thus far described in other streptococci that are arranged in an operon-like cluster with the *comC/DE* genes being transcribed from a single promoter. b) The *comC* gene is separated from the *comD* gene by 148 nucleotides.

**Streptococcus mutans
ComCDE Operon**



Figure 2

Sequences of the open reading frames encoding the signal peptide precursor (ComC), the histidine kinase (ComD), and the response regulator (ComE).

> *S. mutans* comC gene

Encodes a precursor to a signal peptide

[ATGAAAAAAACACTATCATTAAAAAATGACTTTTAAAGAAATTAAGACTGATGAATTAG
AGATTATCATTGGCGGA (AGCGGAAGCCTATCAACATTTTCCGGCTGTTTAACAGAAG
TTTACACAAGCTTTGGGAAAA)] TAA

> *S. mutans* CSP encoding sequence

Competence Signal Peptide

AGCGGAAGCCTATCAACATTTTCCGGCTGTTTAACAGAAGTTTACACAAGCTTTGGG
AAAA [SEQ ID NO:1]

> *S. mutans* comD gene

Encodes a protein that functions as a histidine kinase
receptor

[ATGAATGAAGCCTTAATGATACTTTCAAATGGTTTATTAACTTATCTAACCGTTCTAT
TTCTCTTGTTTCTATTTTCTAAGGTAAGTAATGTCACTTTATCGAAAAAGGAATTA
CTTTTTTCGATAAGCAATTTTCTGATAATGATTGCTGTTACGATGGTGAACGTAAACCT
GTTTTATCCTGCAGAGCCTCTTTATTTTATAGCTTTATCAATTTATCTTAATAGACAGA
ATAGTCTTTCTCTAAATATATTTTATGGTCTGCTGCCTGTTGCCAGTTCTGACTTGTTT
AGGCGGGCAATCATATTTCTTTATCTTGGAATGGAATCAAGGAATTGTAATGGGCAGTAG
CATTATAACCACCTATATGATCGAGTTTGCAGGAATAGCGCTAAGTTACCTCTTTCTCA
GTGTGTTCAATGTTGATATTGGTCGACTTAAAGATAGTTTGACCAAGATGAAGGTCAAA
AAACGCTTGATTCCAATGAATATTACTATGCTTCTATACTACCTTTTAATACAGGTATT
GTATGTTATAGAGAGTTATAATGTGATACCGACTTTAAAAATTTTCGTAAATTTGTCGTTA
TTGTCTATCTTATTTTATTTTGTATTCTGATCTCATTTTTAAGCCAATATACCAAACAA
AAGGTTCAAATGAGATAATGGCACAAGGAAGCTCAGATTGCAAAATATCACCCAGTA
TAGTCAGCAAATAGAATCTCTTTACAAGGATATTTCGAAGTTTCCGCCATGATTATCTGA
ATATTTTAACTAGCCTCAGATTAGGCATTGAAAATAAAGATTTAGCTAGTATTGAAAAG
ATTTACCATCAAATCTTAGAAAAAACAGGACATCAATTGCAGGATACCCGTTATAATAT
CGGCCATCTAGCTAATATTCAAACGATGCTGTCAAGGGTATCTTGTCAGCAAAAATCT
TAGAAGCTCAGAATAAAAAGATTGCTGTCAATGTAGAAGTCTCAAGTAAATAACAACG
CCTGAGATGGAGTTGCTTGATTTTATTACCATCTTTCTATCTTGTTGATAATGCCAT
TGAGGCTGCTTTTCAATCATTAAATCCTGAAATTCAGTTAGCCTTTTTTAAGAAAAATG
GCAGTATAGTCTTTATCATTGAGAATTCACCAAGAAAAACAAATAGATGTGAGTAAA
ATTTTTAAAGAAAATATTCCACTAAAGGCTCCAATCGCGGTATTGGTTTAGCAAAGGT
GAATCATATTTCTGAACATTATCCCAAACAGTTTACAAACAAGCAATCATCATCATT
TATTCAAGCAACTCCTAATAATAAAA] TAG

> *S. mutans* comE gene

Encodes a response regulator that activates transcription
of a number of genes

09533047-041004

[ATGATTTCTATTTTTGTATTGGAAGATGATTTTTTACAACAAGGACGTCTTGAAACCA
CCATTGCAGCTATCATGAAAGAAAAAATTGGTCTTATAAAGAATTGACTATTTTTTGGA
AAACCACAACAACCTTATTGACGCTATCCCTGAAAAGGGCAATCACCAGATTTTCTTTTT
GGATATTGAAATCAAAAAAGAGGAAAAGAAAGGACTGGAAGTAGCCAATCAGATTAGAC
AGCATAATCCTAGTGCAGTTATTGTCTTTGTCACGACACATTCTGAGTTTATGCCCTC
ACTTTTCAGTATCAGGTATCTGCTTTGGATTTTATTGATAAATCTTGAATCCTGAGGA
GTTCTCCACCGCATTGAATCAGCGCTGTATTATGCTATGGAAAACAGCCAGAAGAATG
GTCAATCAGAGGAACTTTTATTTTCCATTCATCTGAAACTCAGTTTCAGGTCCCTTTT
GCTGAGATTCTGTATTTTGAACATCTTCAACAGCCCATAAGCTCTGCCTTTATACTTA
TGATGAACGGATTGAATTCTACGGCAGTATGACTGACATTGTTAAAAATGGATAAGAGAC
TTTTTCAGTGCCATCGCTCTTTTATTGTCAATCCTGCCAATATTACCGTATTGATCGG
AAAAAACGCTTGGCCTATTTTCGAAATAATAAGTCTTGTCTTATTTTCACGAACTAAGTT
AACAAAACCTGAGAGCTGTGATTGCTGATCAAAGGAGAGCAAAA] TGA

1093301-0400

Figure 3

The amino acid sequences of the signal peptide precursor (ComC), the histidine kinase (ComD), and the response regulator (ComE).

> *S. mutans* ComC protein (CSP Precursor)

MKKTLSLKNDFKBIKTDELEIIIGGSGSLSTFFRLFNRSFTQALGK

> *S. mutans* ComD protein (Histidine Kinase)

MNEALMILSNGLLTLYLTVLFLFLFSKVSNTLSKKELTLFSISNFLIMIAVTMVNVNL
FYPAEPLYFIALSIYLNQRNSLSLNIFYGLLPVASSDLFRRATIFFILDGTQGIVMGSS
IITTYMIEFAGIALSYLFLSVFNVDIGRLKDSLTKMKVKKRLIPMNITMLLYYLLIQVL
YVIESYNVIPTLKFRKFVVIVYLILFLILISFLSQYTKQKVQNEIMAQKEAQIRNITQY
SQQIESLYKDIRSFRHDYLNILTSRLGIENKDLASIEKIYHQILEKTGHQLODTRYNI
GHLANIQNDAVKGILSAKILEAQNKKIAVNVEVSSKIQLPEMELLDFFITILSILCDNAI
EAAFESLNPEIQLAFFKKNGSIVFIIONSTKEKQIDVSKIFKENYSTKGSNRGIGLAKV
NHILEHYPKTSLQTSNHHHLFKQLLIK

> *S. mutans* ComE protein (Response Regulator)

MISIFVLEDDFLQQGRLETTIAAIMKEKNWSYKELTIFGKPQQLIDAIPEKGNHQIFFL
DIEIKKEEKKGLEVANQIRQHNPSAVIVFVTTTHSEFMPLTFQYQVSALDFIDKSLNPEE
FSHRIESALYYAMENSQKNGQSEELFIFHSSETQFQVPFAEILYFETSSTAHLCLYTY
DERIEFYGSMTDIVKMDKRLFQCHRSFIVNPANITRIDRKKRLAYFRNKSCLISRTKL
TKLRAVIADQRRAK

1889-00401-0004

The deduced amino acid sequence of the signal peptide precursor in various strains and its predicted cleavage site. The original peptide is expressed as a 46-amino acid peptide that is cleaved after the glycine-glycine residues to generate an active signal peptide.

BM71 CSP	1	MKKTPSLKNDFKFKIKTDELEIIIGGSGSLSTFFRLFNRSFTQALGK	46
GB14 CSP	1	MKKTILSLKNDFKSIKTDELEIIIGGSGSLSTFFRLFNRSFTQALGK	46
H7 CSP	1	MKKTILSLKNDFKFKIKTDELEIIIGGSGSLSTFFRLFNRSFTQALGK	46
JH1005 CSP	1	MKKTILSLKNDFKFKIKTDELEIIIGSGTSLSTFFRLFNRSFTQA	43
LT11 CSP	1	MKKTILSLKNDFKFKIKTDELEIIIGGSGSLSTFFRLFNRSFTQALGK	46
NG8 CSP	1	MKKTILSLKNDFKFKIKTDELEIIIGGSGSLSTFFRLFNRSFTQALGK	46
UAB159 CSP	1	MKKTILSLKNDFKFKIKTDELEIIIGGSGSLSTFFRLFNRSFTQALGK	46

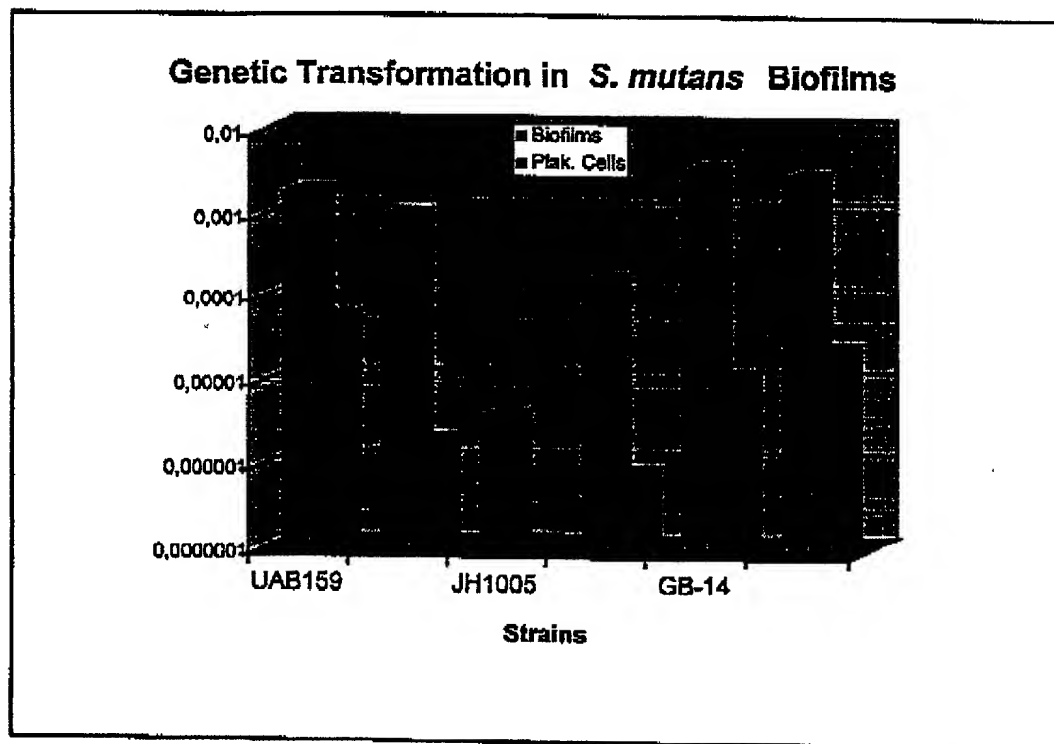
consensus: 1 MKKTL¹SLKND²FKEIKTDELEIIIGG SGSLSTFFRLFNRSPTQALGK 46
predicted cleavage site: ^

The synthetic signal peptide that is effective at inducing competence, biofilm formation and acid tolerance in *Streptococcus mutans*.

SGSLSTFFRLFNRSFTQALGK [SEQ ID NO:2]

Figure 6

The natural activity of the signal/receptor system functioning *in vitro* in model biofilms as determined by the ability of various strains of *S. mutans* to accept donor plasmid DNA conferring erythromycin resistance.



Strain	Peptide added Number of Transformants/Recipients	No peptide Number of Transformants/Recipients
UAB15	4.65×10^{-1}	1.78×10^{-6}
JH1005 ²	6.98×10^{-2}	0

The strain contains a nonsense mutation in the *comC* gene encoding the CSP.

Figure 8

List of the primers used to amplify the genes or internal regions of the target genes by polymerase chain reaction (PCR) for subsequent sequencing or inactivation.

ComC region

ComC Primer Pair: F5-B5

[F5] 23406-23424 5'- AGTTTTTTGTCTGGCTGCG -3'

19 nt forward primer

pct G+C: 47.4 Tm: 50.5

[B5] 24056-24037 5'- TCCACTAAAGGCTCCAATCG -3'

20 nt backward primer

pct G+C: 50.0 Tm: 51.9

651 nt product for F5-B5 pair (23406-24056)

Optimal annealing temp: 50.3

pct G+C: 30.9 Tm: 71.5

ComD region

ComD Primer Pair: F1-B1

[F1] 392-415 5'- CGCTAAGTTACCTCTTTCTCAGTG -3'

24 nt forward primer

pct G+C: 45.8 Tm: 51.6

[B1] 683-663 5'- GCTTCCTTTTGTGCCATTATC -3'

21 nt backward primer

pct G+C: 42.9 Tm: 50.8

292 nt product for F1-B1 pair (392-683)

Optimal annealing temp: 49.5

pct G+C: 30.8 Tm: 70.2

ComE region

ComE Primer Pair: F1-B1

[F1] 145-165 5'- CCTGAAAAGGGCAATCACCAG -3'

21 nt forward primer

pct G+C: 52.4 Tm: 55.9

[B1] 606-585 5'- GCGATGGCACTGAAAAAGTCTC -3'

22 nt backward primer

pct G+C: 50.0 Tm: 55.4

462 nt product for F1-B1 pair (145-606)

Optimal annealing temp: 53.6

pct G+C: 38.3 Tm: 74.1

093301-041001

Figure 9

ComCDE local region. The ComC (first highlighted region; nucleotides 101 to 241), ComD (second highlighted region; nucleotides 383 to 1708) and ComE (third highlighted region; nucleotides 1705 to 2457) proteins are highlighted.

Sequence Range: 1 to 2557

```

      10      20      30      40      50
ACATTATGTGTCCTAAGGAAATATTACITTTTCAAGAAATCCATGATT
TGTAATACACAGGATTCCTTTTATAATGAAAAAGTTCTTTTAGGTACTAA
                                <K K L F I W S K
                                <
      60      70      80      90     100
TTTTCATAAAAATAGTATACTAATTATAATCAAAAAAGGAGATATAAA
AAAAGTATTTTATCATATGATTAATATTAGTTTTCCTCTATATTT
<K M F F L I S I I I L F L L Y L
<
     110     120     130     140     150
ATGAAAAAAACACTATCATTAAAAAATGACTTTAAAGAAATTAAGACTGA
TACTTTTTTGTGATAGTAATTTTACTGAAATTTCTTAAATCTGACT
M K K T L S L K N D F K E L K T D
                                ORF RF [2]
<I F F V S D N F F S K L S I L V S
<
     160     170     180     190     200
TGAATTAGAGATTATCATTGGCGGAAGCGGAAGCCTATCAACATTTTCC
ACTTAATCTCTAATAGTAACCGCCTTCGCCTTCGGATAGTTGTAAAAAGG
E L E I I I G G S G S E S T F E
                                ORF RF [2]
<S N S I I M
<
     210     220     230     240     250
GGCTGTTTAAACAGAAGTTTACACAAGCTTTGGGAAAAATAAGATAGGCTA
CCGACAAATTGTCTTCAAAATGTGTTCGAAACCCTTTATTCTATCCGAT
E L F N R S F T Q A L G A K
                                ORF RF [2]
                                >
     260     270     280     290     300
ACATIGGAATAAAACAAGGCTGGATTTATTATTCAGCCTTTTAAATGT
TGTAACCTTATTTTGTTCGACCTAAATAATAAGGTCGAAAAATTTACA
     310     320     330     340     350
AAAAATAAAATACAGGGTTAAATAATCAAGTGTGCTGCTGGATGAGAA
TTTTATTTTATGTCCCAATTTATTAGTTTCACGACAGCACCTACTCTT
     360     370     380     390     400
GATAAACTATCTCTTAGAGAATAGGCCTCCTCTATTTTATTATTAGGAG
CTATTTTGATAGAGAATCTCTTATCCGAGGAGATAAAATAAATCCTC
                                <K I L L L
                                < ORF RF [
     410     420     430     440     450
```

0933017 041001

AACGAACTTATTTACTACTACTAACGAACAAACATTTGACCAAAACCCCTA
O K E L H E E N S L L O L S T K E V
ORF RF[4] C

460 470 480 490 500
AATGTTCAAGAATATGATTACCTTTGCTAAACCAATACCGCGATTGGAG
TTACAAGTTCTTATACTAAGTGGAAACGATTTGGTTATGGCGCTAACCTC
H E L I H N V K A L G I G R N S
ORF RF[4] C

510 520 530 540 550
CCTTTAGTGGGAATAGTTTTCTTTAAAAATTTACTCACATCTATTGTTT
GGAAATCACCTTATCAAACAAATTTTTAAAAATGAGTGTAGATAAACAA
G K T S Y N E K F I K S V D T O K
ORF RF[4] C

560 570 580 590 600
TTCTTTGGTGGGAATTCGAATGATAAAGACTATACTGCCATTTTTCTTAA
AAGAAACCACCTTAAGACTTACTATTCTGATATGACGGTAAAAAGAATT
E K T S N O T I F V I S G N K K F
ORF RF[4] C

610 620 630 640 650
AAAAGGCTAACTGAATTTCAAGATTTAATGATTGAAAGCAGCCTCAATG
TTTTCCGATTGACTTAAAGTCCTAAATTACTAAGCTTTCGTGGGAGTTAC
M>

F A L O I E P N L S E F A R E I
ORF RF[4] C

660 670 680 690 700
GCATTATCACACAAGATAGAAAGTATGGTAATGAAATCAAGCAACTCCAT
CGTAATAGTGTGTTCTATCTTTCATACCATTACTTTAGTTCGTTGAGGTA
A L S H K I E S M V M K S S N S I>
ORF RF[3]
A N D C E I S L I T I F D L E E M
ORF RF[4] C

710 720 730 740 750
CTCAGGCAGTGTATTTTACTTGAGACTTCTACATTGACAGCAATCTTTT
GAGTCCGTCAACATAAAATGAACTCTGAAGATGTAAGTGTGTTAGAAAA
S G S C I L L E T S T L T A I F>
ORF RF[3]
E P L O I K S S V E V N V A I K K
ORF RF[4] C

760 770 780 790 800
TATTCTGAGCTTCTAAGATTTTCTGCTGACAAGATACCTTGACAGCATCG
ATAAGACTCGAAGATTCTAAAAACGACTGTTCTATGGGAAGTGTGTTAGC
L F>

N Q A E L I K A S L I G K V A D
ORF RF[4] C

810 820 830 840 850
TTTTGAATATTAGCTAGATGCGCGATATTATAACGGGTATCCTGCAATTG
AAAACTTATAATCGATCTACCGGCTATAATATTGCCCATAGGACGTTAAC

09833017104006

< N Q L N A L H S I N Y R P D Q L Q
ORF RF[4] C

860 870 880 890 900
ATGTCCTGTTTTCTAAGATTGATGGTAAATCTTTCAATACTAGCTA
TACAGGACAAAAAGATTCTAAACTACCATTAGAAAAGTTATGATCGAT
< H G T K E L I O R Y L K E T S A L
ORF RF[4] C

910 920 930 940 950
AATCTTTATTTCAATGCCTAATCTGAGGCTAGTTAAAAATATTCAGATAA
TTAGAAATAAAAGTTACGGATTAGACTCCGATCAATTTTATAAGTCTATT
< D K N E L G L R L S T L I N L Y
ORF RF[4] C

960 970 980 990 1000
TCATGGCGAAACTTCGAATATCCTTGTAAAGAGATTCTATTTGCTGACT
AGTACCGCCTTTGAAGCTTATAGGAACATTTCTCTAAGATAAACGACTGA
M A E T S N I L V K R F Y L L T >
< D H R E S R I D K Y L S E L Q Q S
ORF RF[4] C

1010 1020 1030 1040 1050
ATACTGGGTGATATTTTGAATCTGAGCTTCTTTTGTGCCATTATCTCAT
TATGACCCACTATAAAGCTTAGACTCGAAGGAAAACACGGTAATAGAGTA
I L G D I S N L S F L L C H Y L I >
< V O T I N R I O A E K Q A M T E N
ORF RF[4] C

1060 1070 1080 1090 1100
TTTGAACCTTTTGTGTTGTTATATTGGCTTAAAAATGAGATCAGAATCAAA
AACTTGGAAAAACAACCATATAACCGAATTTTACTCTAGTCTTAGTTT
L N L L F G I L A >
< Q V K Q K T Y Q S L F S L F I L
ORF RF[4] C

1110 1120 1130 1140 1150
AATAAAATAAGATAGACAATAACGACAAATTTACGAAATTTTAAAGTCGG
TTATTTTATTCTATCTGTTATTGCTGTTTAAATGCTTAAATTTTCAGCC
< F L I L Y V E V V E K R F K L T P
ORF RF[4] C

1160 1170 1180 1190 1200
TATCACATTATAACTCTCTATAACATACAATACCTGTATTAAAAGGTAGT
ATAGTGTAATATTGAGAGATATTGTATGTTATGGACATAATTTTCCATCA
< I V N Y S E I V Y L V O I L L Y Y
ORF RF[4] C

1210 1220 1230 1240 1250
ATAGAAGCATAGTAATATTTCATTGGAATCAAGCGTTTTTTGACCTTCATC
TATCTTCGTATCATTATAAGTAACCTTAGTTCGCAAAAACTGGAAGTAG
< L L M T I N M E L F K K V K M
ORF RF[4] C

09533047 044004

1260 1270 1280 1290 1300
TTGGTCAAACTATCTTTAAGTCGACCAATATCAACATTGAACACACTGAG
AACCAGTTTGATAGAAATTTCAGCTGGTTATAGTTGTAACCTGTGTGACTC
<K P L S D K L R G E D V N P V S L
< ORF RF[4] C

1310 1320 1330 1340 1350
AAAGAGGTAACCTTAGCGCTATTCTCCTGCAAACTCGATCATATAGGTGGTTA
TTTCTCCATTGAATCGCGATAAGGACGTTTGAGCTAGTATATCCACCAAT
<P L Y S L A I G A F E I M Y E T I
< ORF RF[4] C

1360 1370 1380 1390 1400
TAATGCTACTGCCCATTACAATTCCTTGAGTTCCATCCAAGATAAAGAAT
ATTACGATGACCGGTAATGTTAAGGAACTCAAGGTAGGTTCTATTTCTTA
<I S S G M V I G Q T G D L I P F
< ORF RF[4] C
<L E K L E M W S L S Y
<

1410 1420 1430 1440 1450
ATGATTGCCCGCCTAAACAAGTCAGAACTGGCAACAGGCAGCAGACCATA
TACTAACCGCGCGGATTGTTCAGTCTTGACCGTTGTCCGTCGTCGGTAT
<I T A P R E F L D S S A V P T L I G Y
< ORF RF[4] C
<S Q G G L C T L V P L L C C V M
<

1460 1470 1480 1490 1500
AAATATATTTAGAGAAAGACTATTCTGTCTATTAAGATAAATTGATAAAG
TTTATATAAATCTCTTTCTGATAAGACAGATAATTCTATTTAACTATTTT
<F I N L S L S N O R N L Y L S L A
< ORF RF[4] C

1510 1520 1530 1540 1550
CTATAAAATAAAGAGGCTCTGCAGGATAAAACAGGTTTACGTTCCACCATC
GATATTTTATTTCTCCGAGACGTCCTATTTTGTCCAAATGCAAGTGGTAG
<I F Y L P E A P Y F L N Y N V M
< ORF RF[4] C

1560 1570 1580 1590 1600
GTAACAGCAATCATTATCAGAAAATTGCTTATCGAAAAAAGAGTTAATTC
CATTGTGCTTAGTAATAGTCTTTTAAACGAATAGCTTTTCTCAATTAAG
<T V A I N I L F N S I S P I L T E E
< ORF RF[4] C

1610 1620 1630 1640 1650
CTTTTTCGATAAAGTGACATTACTTACCTTAGAAAAATAGAAACAAGAGAA
GAAAAAGCTATTTCACTGTAATGAATGGAATCTTTATCTTTGTTCTCTT
<K K S L T V N S V K S F L F L L F
< ORF RF[4] C

1660 1670 1680 1690 1700
ATAGAACGGTTAGATAAGTTAATAAACCATTGAAAGTATCATTAAAGGCT
TATCTTGCCAATCTATTCAATTATTTGGTAACTTTCATAGTAATCCGA
<I V T E Y T L L G N S L I M L A
< ORF RF[4] C

0033017-041001

1710 1720 1730 1740 1750
TCATTCATTTTGGCTCTCTCTTTGATCAGCAATCACAGCTCTCAGTTTGT
AGTAAGTAAACGAGAGGAACTAGTCGTTAGTGTGAGAGTCAAAACAA
<E N M

<K A R R Q D A I V A R L K T
< ORF RF[5] C

1760 1770 1780 1790 1800
AACTTAGTTCGTGAAATAAGACAAGACTTATTATTTGAAAATAGGCCAA
TTGAATCAAGCACTTTTATTCTGTTCTGAATAATAAGCTTTTATCCGCTT
<L K L R S L L C S K N N R F Y A E
< ORF RF[5] C

1810 1820 1830 1840 1850
GCGTTTTTTCGATCAATACGGGTAATATTGGCAGGATTGACAATAAAG
CGCAAAAAGGCTAGTTATGCCATTATAACCGTCCTAACTGTTATTTTC
<R K K R D L R T L N A F N V I F S
< ORF RF[5] C

1860 1870 1880 1890 1900
AGCGATGGCACTGAAAAAGTCTCTTATCCATTTTAACAATGTCAGTCATA
TCGCTACCGTGACTTTTTCAGAGAATAGGTAAAATTGTTACAGTCAGTAT
M A L K K S L I H F N N V S H>
ORF RF[1]
<R H C Q F L R K D M K V I D T M
< ORF RF[5] C

<V
<

1910 1920 1930 1940 1950
CTGCCGTAGAATTCAATCCGTTTCATCATAAGTATAAAGGCAGAGCTTATG
GACGGCATCTTAAGTTAGGCAAGTAGTATTTCATTTCCGTCCTCGAATAC
T A V E F N P F I I S I K A E L M>
ORF RF[1]
<S G Y F E I R E D Y T Y L C L K H
< ORF RF[5] C
<A T S N L G N M M L I F A S S I
< ORF RF[6] C

1960 1970 1980 1990 2000
GGCTGTGAAGATGTTTCAAAATACAGAATCTCAGCAAAAGGGACCTGAA
CCGACAACTTCTACAAAGTTTATGTCTTAGAGTCGTTTTCCCTGGACTT
G C>

<A T S S T E F Y L E E A F P V O F
< ORF RF[5] C
<P Q Q L H K L I C F R L L L S R F
< ORF RF[6] C

2010 2020 2030 2040 2050
ACTGAGTTTCAGATGAATGGAATAAAGTTTCCTCTGATTGACCATTC
TGACTCAAAGTCTACTTACCTTTTATTTTCAAGGAGACTAACTGGTAAG
<Q T E S S H F I F L E E S Q G N
< ORF RF[5] C
<S L K L H I S F L F N R Q N V M R

FOOTNOT 04007

< _____ ORF RF[6] C _____

2060 2070 2080 2090 2100
TTCTGGCTGTTTTCCATAGCATAATAACAGCGCTGATTCAATGCGGTGGGA
AAGACCGACAAAAGGTATCGTATTATGTCCGCGACTAAGTTACGCCACCCCT
<K O S N E M A Y Y L A S E I F E S
< _____ ORF RF[5] C _____

<R A T K W L M
< _____ ORF RF[6] C _____

2110 2120 2130 2140 2150
GAACTCCTCAGGATTCAAAGATTTATCAATAAAATCCAAAGCAGATACCT
CTTGAGGAGTCCTAAGTTTCTAAATAGTTATTTIAGGTTTCGTCTATGGA
<F E E P N L S K D I F E L A S V Q
< _____ ORF RF[5] C _____

2160 2170 2180 2190 2200
GATACTGAAAAGTGAGGGGCATAAACTCAGAAATGTGTGCTGACAAAGACA
CTATGACTTTTCACTCCCCGTATTTGAGTCTTACACAGCACTGTTTCTGT
M C R D K D>

<Y O F T L E M F E S H T P V P V
< _____ ORF RF[5] C _____

2210 2220 2230 2240 2250
ATAACTGCACTAGGATTATGCTGTCTAATCTGATTGGCTACTTCCAGTCC
TATTGACGTGATCCTAATACGACAGATTAGACTAACCGATGAAGGTCAGG
N N C T R I M L S N L I G Y F Q S>

<I V A S E N H O R I O N A V E L G
< _____ ORF RF[5] C _____

2260 2270 2280 2290 2300
TTTCTTTTCTCTTTTTTTGATTTCATATCCAAAAAGAAAATCTGGTGAT
AAAGAAAAGCAGAAAAAACTAAAGTTATAGGTTTTCTTTTAGACCACTA
F L F L F F D F N I Q K E N L V I>

<K K E E K K I E I D L F E I O H N
< _____ ORF RF[5] C _____

2310 2320 2330 2340 2350
TGCCCTTTTCAGGGATAGCGTCAATAAGTTGTTGTGGTTTTTCCAAAAATA
ACGGGAAAAGTCCCTATCGCAGTTATTCAACAACACCAAAGGTTTTTAT
A L F R D S V N K L L W F S K N>

<G K E P I A D T L Q O P K G F I
< _____ ORF RF[5] C _____

2360 2370 2380 2390 2400
GTCAATTCTTTATAAGACCAATTTTTTTCTTTTCATGATAGCTGCAATCGT
CAGTTAAGAAATATTCTGGTTAAAAAAGAAAGTACTATCGACGTTACCA
S Q F F I R P I F F F H D S C N G>

M I A A M V>

<T L E K Y S N N K E K M I A A I T
< _____ ORF RF[5] C _____

09633017/04001

2410 2420 2430 2440 2450
GGTTTCAGACGTCCTTGTTGTAAAAAATCATCTTCCAATACAAAAATAG
CCAAAGTTCTGCAGGAACAACATTTTTAGTAGAAGGTTATGTTTTTATC
G F K T S L L>

V S R R P C C K K S S S N T K I>

<T E G F G Q Q L F D D E L V E I S
ORF RF[5] C

2460 2470 2480 2490 2500
AAATCATTATTTCTCCTTTAATCTTCTATTTAGGTTAGCTGATTAACACT
TTTAGTAATAAAGAGCAAATTAGAAGATAAATCCAATCGACTAATTGTGA
E I I I S P L I F Y L G>

<T E G F G Q Q L F D D E L V E I S

2510 2520 2530 2540 2550
ATACACAGAAAAGGTATAAAACGATATCACTCAATAAAATCTACTAACTT
TATGTGTCCTTTCCATATTTTGCTATAGTGAGTTATTTTACATGATTGAA

AATAACC
TTATTGG

0933037-043004

Figure 10

The comX nucleotide sequence, amino acid sequence, and its local region with 100bp included both upstream and downstream (promoter is upstream).

> *S. mutans* comX gene

ATGGAAGAAGATTTTGAATTTGTTTTTAATAAGGTTAAGCCAATTGTATGGAAATTAAG
CCGTTATTACTTTTATTAAATGTGGACTCGTGAAGATTGGCAACAAGAGGGAATGTTGA
TTTTGCACCAATTATTAAGGGAACATCCAGAATTAGAAGAGGATGATACAAAATTGTAT
ATCTATTTTAAGACACGTTTTTCTAATTACATTAAAGATGTTTTGCGTCAGCAAGAAAG
TCAGAAACGTCGTTTTAATAGAATGTCTTATGAAGAAGTCGGTGAGATTGAACACTGTT
TGTCAAGTGGCGGTATGCAATTGGATGAATATATTTTATTTTCGTGATAGTTTGCTTGCA
TATAACAAGGTCTGAGTACTGAAAAGCAAGAGCTGTTTGAGCGCTTGGTAGCAGGAGA
GCACTTTTTGGGAAGGCAAAGTATGCTGAAAGATTTACGTAAAAAATTAAGTGATTTTA
AGGAAAAA

> *S. mutans* ComX protein

MEEDFEIVFNKVKPIVWKLRSYYFIKMWTRWDWQQEGLILHQLLREHPELEEDDTKLY
IYFKTRFSNYIKDVLRRQESQKRRFNRMSEYEEVGEIEHCLSSGGMQLDEYILFRDSL
LLAYKQGLSTEKQELFERLVAGEHFLGRQSMKDLRKKLSDFKEK

> *S. mutans* comX gene local region

GTAAATAAAACAGCCAGTTAAGATGGGACATTTATGTCCTGTTCTTAAAGTCTTTTTCG
TTTTATAATAATTTTATTATAAAAGGAGGTCATCGTAATAGATGGAAGAAGATTTTGAA
ATTGTTTTTAATAAGGTTAAGCCAATTGTATGGAAATTAAGCCGTTATTACTTTATTAA
AATGTGGACTCGTGAAGATTGGCAACAAGAGGGAATGTTGATTTTGCACCAATTATTAA
GGGAACATCCAGAATTAGAAGAGGATGATACAAAATTGTATATCTATTTTAAGACACGT
TTTTCTAATTACATTAAAGATGTTTTGCGTCAGCAAGAAAGTCAGAAACGTCGTTTTAA
TAGAATGTCTTATGAAGAAGTCGGTGAGATTGAACACTGTTTGTCAAGTGGCGGTATGC
AATTGGATGAATATATTTTATTTTCGTGATAGTTTGCTTGCAATATAACAAGGTCTGAGT
ACTGAAAAGCAAGAGCTGTTTGAGCGCTTGGTAGCAGCAGAGCACTTTTTGGGAAGGCA
AAGTATGCTGAAAGATTTACGTAAAAAATTAAGTGATTTTAAGGAAAAATAGTTAAAA
GGGAAGAATGGAACATGTGATTGTACCATTCTTTTTGGTTGAAAATTAAGAAAAGTTA
TTATAAATTATTGGTTTAACATGCCATATTA

FIG. 10

Figure 11.

The comA and comB nucleotide and amino acid sequences. ComA and ComB are the components of the CSP exporter.

➤ *S. mutans* comA gene

ATGAAACAAGTTATTTATGTTGTTTTAATCGTCATAGCCGTTAACATTCTCTTAGAGAT
TATCAAAAGAGTAACAAAAGGGGAGGGACAGTTTCGTCATCTAATCCTTTACCAGATG
GGCAGTCTAAGTTGTTTTGGCGCAGACATTATAAGCTAGTACCTCAGATTGATACCAGA
GACTGTGGGCCCGGCAGTGTCTGGCATCTGTTGCAAAGCATTACGGATCTAATTACTCTAT
CGCTTATCTGCGGGAACCTCTCAAAGACTAACAAGCAGGGAACAACAGCTCTTGGCATTG
TTGAAGCTGCTAAAAAGTTAGGCTTTGAAACACGCTCTATCAAGGCGGATATGACGCTT
TTTGATTATAATGATTTGACCTATCCTTTTATCGTCCATGTGATTAAAGGAAAACGTCT
GCAGCATTATTATGTCGTCTATGGCAGCCAGAATAATCAGCTGATTATTGGAGATCCTG
ATCCTTCAGTTAAGGTGACTAGGATGAGTAAGGAACGCTTTCAATCAGAGTGACAGGC
CTTGCAATTTTCTAGCTCCTCAGCCTAACTATAAGCCTCATAAAGGTGAAAAAATGG
TTTGTCTAATTTCTTCCCGTTGATCTTTAAGCAGAAAGCTTTGATGACTTATATTATCA
TAGCTAGCTTGATTGTGACGCTCATTGATATTGTGCGGATCATACTATCTCCAAGGAATA
TTGGACGAGTACATTCTGATCAGCTGATTTCAACTTTAGGAATGATTACGATTGGTCT
GATAATAACCTATATTATCCAGCAGGTCATGGCTTTTGCAAAGAATACCTCTTGGCCG
TACTCAGTTTGCGTTTAGTCATTGATGTTATCCTGTCTTATATCAAACATATTTTACG
CTTCCTATGTCTTTCTTTGCGACAAGGCGAACAGGAGAAATCACGTCTCGTTTACAGA
TGCCAAATCAGATTATTGATGCTGTAGCGTCAACCATCTTTCAATCTTTTATGATATGA
CTATGGTAATTTTGGTTGGTGGGTTTTGTTGGCGCAAAACAATAACCTTTTCTTTCTA
ACCTTGCTCTCCATTCCGATTTATGCCATCATTATTTTGGCTTTCTTGAAACCTTTGA
GAAAATGAATCACGAAGTGATGGAAGCAATGCTGTGGTAAGTTCTTCTATCATTGAAG
ATATCAATGGGATCGAAACCATTAATCACTCACAAGTGAGTCCGCTCGTTATCAAAC
ATTGATAGTGAATTTGTTGATTATTTGGAGAAAACTTTAAGCTACACAAGTATAGTGC
CATTCAAACCGCATTAAAAAGCGGTGCTAAGCTTATCCTCAATGTTGTCATTCTCTGGT
ATGGCTCTCGTCTAGTTATGGATAATAAAATCTCAGTTGGTCAGCTTATCACCTTTAAT
GCTTTGCTGTCTTATTTCTCAAATCCAATTGAAAATATTATCAATCTGCAATCCAACT
GCAGTCAGCTCGCGTTGCCAATACACGTCTTAATGAGGTCTATCTTGTCGAATCTGAAT
TTGAAAAAGACGGCGATTTATCAGAAAATAGCTTTTTAGATGGTGATATTTCTGTTGAA
AATCTTTCTTATAAATATGGATTGCGGCGAGATACCTTATCAGATATTAATTTATCAAT
CAAAAAGGCTCCAAGGTCAGTCTAGTTGGAGCCAGTGGTCTGGTAAACAACCTTGG
CTAAACTGATTGTCAATTTCTACGAGCCTAACAAGGGGATTGTTCAATCAATGGCAAT
GATTTAAAAGTTATTGATAAGACAGCTTTGCGGCGGCATATTAGCTATTTGCCGCAACA
GGCCTATGTTTTTAGTGGCTCTATTATGGATAATCTCGTTTTAGGAGCTAAGAAGGAA
CGAGTCAGGAAGACATTATTCGTGCTTGTGAAATTGCTGAAATCCGCTCGGACATTGAA
CAAATGCCCTCAGGGCTATCAGACAGAGTTATCAGATGGTGCCGGTATTTCTGGCGGTCA
AAAACAGCGGATTGCTTTAGCTAGGGCCTTATTAACACAGGCACCGGTTTTGATTCTGG
ATGAAGCCACCAGCAGTCTTGATATTTTGACAGAAAAGAAAATATCAGCAATCTCTTA
CAGATGACGGAGAAAACAATAATTTTGTGTTGCCACCGCTTAAGCATTTCACAGCGTAC
TGACGAAGTCATTGTCTATGGATCAGGGAAAAATGTTGAACAAGGCACTCATAAGGAAC
TTTATAGCTAAGCAAGGTTTCTATTATAACCTGTTTAAT

➤ *S. mutans* ComA protein

09833017 041001

MKQVIYVVLIVIAVNILLEIIKRVTKRGGTVSSSNPLPDGQSKLFWRRHXYKLVPQIDTR
 DCGPAVLASVAKHYGSNYSIAYLRELSKTNKQGTALGIVEAAKKGFEFTRSIKADMTL
 FDYNDLTYPFIVHVIKGRKLQHYVVGYSQNNQLIIIGDPDPSVKVTRMSKERFQSEWTG
 LAIFLAPQPNYKPHKGEKNGLSNFFPLIFKQKALMTYIIIIASLIVTLIDIVGSYYLQGI
 LDEYIPDQLISTLGMITIGLIITYIIIQQVMAFAKEYLLAVLSLRLVIDVILSYIKHIFT
 LPMSFFATRRTGEITSRFTDANQIIDAVASTIFSIFLDMTMVILVGGVLLAQNNNLFFL
 TLLSIPIYAIIIIFAFCLKPFKMNHEVMESNAVVSSSIIEDINGMETIKSLTSESARYQN
 IDSEFVDYLEKNFKLHKYSAIQTALKSGAKLILNVVILWYGSRLVMDNKISVGQLITFN
 ALLSYFSNPIENIINLQSKLQSARVANTRLNEVYLVESEFEKDGDLSENSFLDGDISFE
 NLSYKYGFGRDTLSDINLSIKKGSKVSILVGASGSGKTTLAKLIVNFYEPNKGIVRINGN
 DLKVIDKTALRRHISYLPQQAYVFGSISMDNLVLCAKEGTSQEDIIRACEIAEIRSDIE
 QMPQGYQTELSDGAGISGGQKQRIALARALLTQAPVLILDEATSSLDILTEKKIISNLL
 QMTEKTIIFVAHRLSISORTDEVIVMDQGKIVEOGTHKELLAKQGFYYNLFN

➤ *S. mutans* comB gene

ATGGATCCTAAATTTTACAAAGTGCAGAATTTTATAGGAGACGCTATCATAATTTTGC
GACACTATTAATTGTTCTTTGGTCTGCTTGATTATCTTCTTGGTCATATTCCTTTGTT
TTGCTAAAAAAGAAATTACAGTGATTTCTACTGGTGAAGTTGCACCAACAAAGGTTGTA
GATGTTATCCAATCTTACAGTGACAGTTCAATCATTAATAATAATTTAGATAATAATGC
AGCTGTTGAGAAGGGAGACGTTTTAATTGAATATTAGAAAATGCCAGTCCAAACCGTC
AGACTGAACAAAAGAATATTATAAAAGAAAGACAAAAACGAGAAGAGAAGGAAAAGAAA
AAACACCAAAAGAGCAAGAAAAAGAAGTCTAAGAGCAAGAAAGCTTCCAAAGATAA
GAAAAAGAAATCGAAAGACAAGGAAAGCAGCTCTGACGATGAAAATGAGACAAAAAAGG
TTTCGATTTTTGCTTCAGAAGATGGTATTATTATACCAATCCCAAATATGATGGTCCC
AAATATTATTCCGAAGCAAACCGAGATTGCTCAAATCTATCCTGATATTCAAAAAACAAG
AAAAGTGTTAATCACCTATTATGCTTCTTCTGATGATGTTGTTTCTATGAAAAAGGGGC
AAACCGCTCGTCTTTCTTGGAAAAAAGGGAAATGACAAGGTTGTTATTGAAGGAAAA
ATTAACAATGTCGCTTCATCAGCAACTACTACTAAAAAAGGAAATCTCTTTAAGGTTAC
TGCCAAAGTAAAGGTTTCTAAGAAAAATAGCAAACTCATCAAGTATGGTATGACAGGCA
AGACAGTCACTGTCAATGATAAAAAAGACTTAATTTTGATTATTTCAAAGATAAATTACTG
CATAAAATGGATAAT

➤ S. mutans ComB protein

MDPKFLQSAEFYRRRYHNFATLLIVPLVCLIIIFLVIFLCFAKKEITVISTGEVAPTQVV
DVIQSYSDSSIIKNLNDNNAAVEKGDVLEIYSENASPNRQTEQKNIIKERQKREEKEKK
KHQKSKKKKKSKSKKASKDKKKKSKDKESSDDENETKKVSI FASEDGI IHTNPKYDGA
NIIPKQTEIAQIYPDIQKTRKVLITYYASSDDVVSMMKKGTARLSLEKKGNDKVVIEGK
INNVAASSATTTKGNLFKVTAQVQVSKNSKLIKYGMTGKTVTVIDKRTYFDYFKDKLL
HKMDN

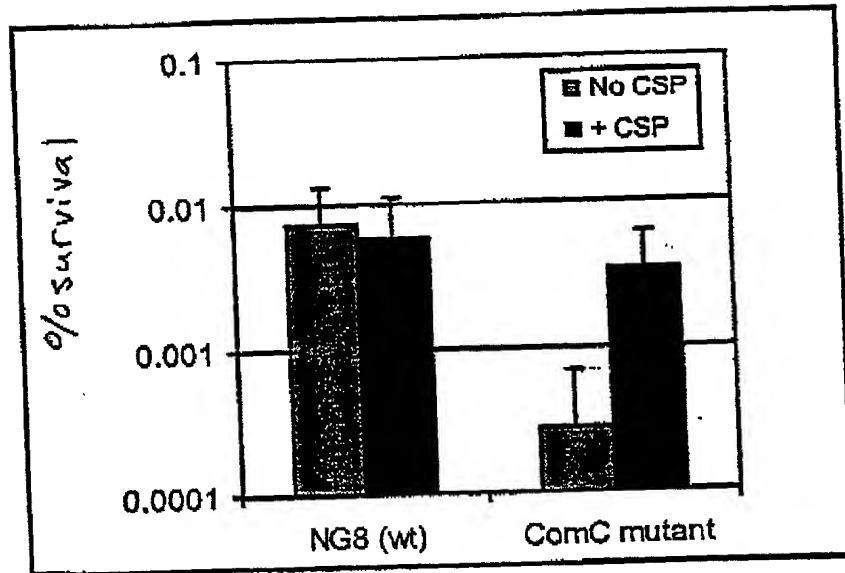


Figure 12